Chapter Seven: Cross Tabulations and Correlations between variables

In this chapter, we'll look at how PASW for Windows can be used to create contingency tables, oftentimes called cross tabulations (or crosstabs), bivariate, or two-variable tables. A contingency table helps us look at whether the value of one variable is associated with, or "contingent" upon, that of another. It is most useful when each variable contains only a few categories. Usually, though not always, such variables will be nominal or ordinal. Some techniques for examining relationships among interval or ratio variables are presented in later chapters.

To make it easier to follow the instructions in this chapter, we recommend that you set certain options in PASW in the same way that are shown here. First, click on 'Edit" in the menu bar, then on "Options," and "General." Under "Variable List," click on "Display names," and "Alphabetical." These choices will ensure that the variables in dialog boxes will look like they do in our examples (see Figure 7-1).

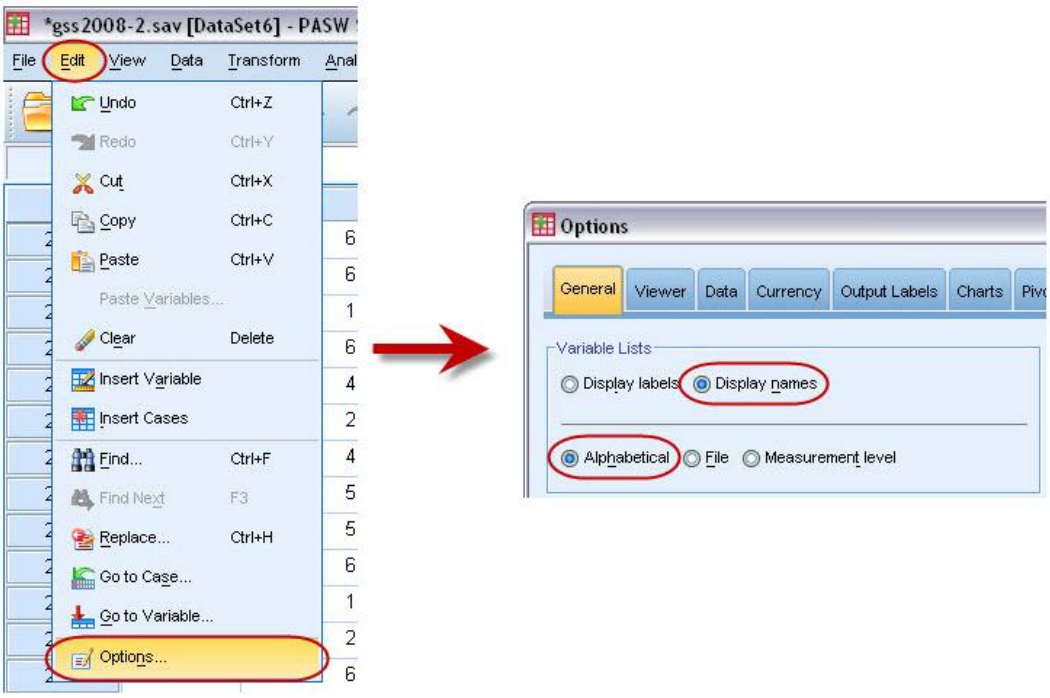

Figure 7- 1

Now click on "Pivot Tables." PASW offers a number of different "looks" for contingency tables. You might want to experiment with the different choices. For now, however, click on "Academic" under "TableLook," and on "Labels and data" under "Adjust Column Widths" (see Figure 7-2). Then click on OK.

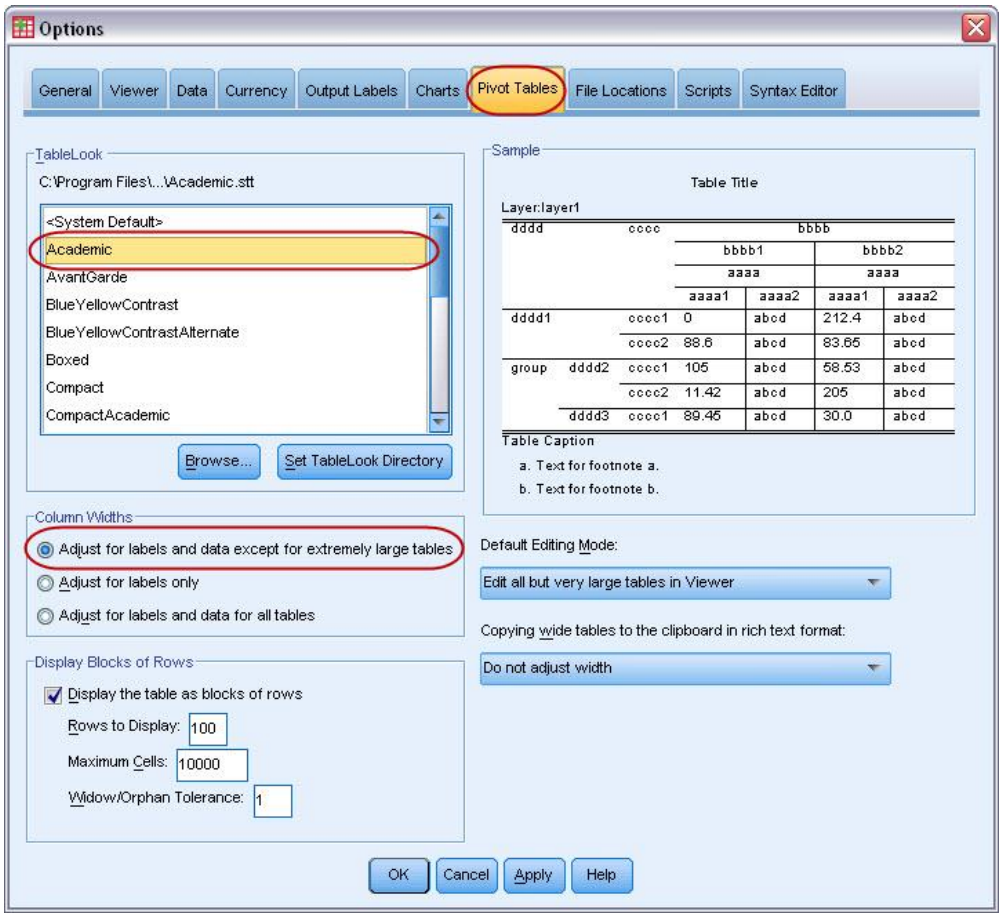

Figure 7- 2

To illustrate the Crosstabs technique, we'll use the General Social Survey subset (GSS2008.sav).

Crosstabs are particularly useful in hypothesis testing. Let's see if there is any difference between men and women in their attitudes towards abortion. SEX is the independent variable, while attitudes towards abortion is the dependent variable. To create a contingency table (crosstabs), from the menu choose "Analyze," "Descriptive Statistics," and "Crosstabs". As seen in the figure 7.3.

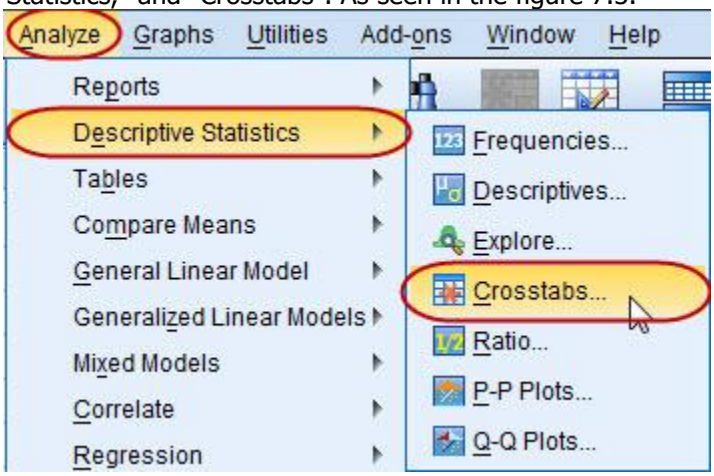

Figure 7- 3

A dialog box opens as shown in Figure 7-4.

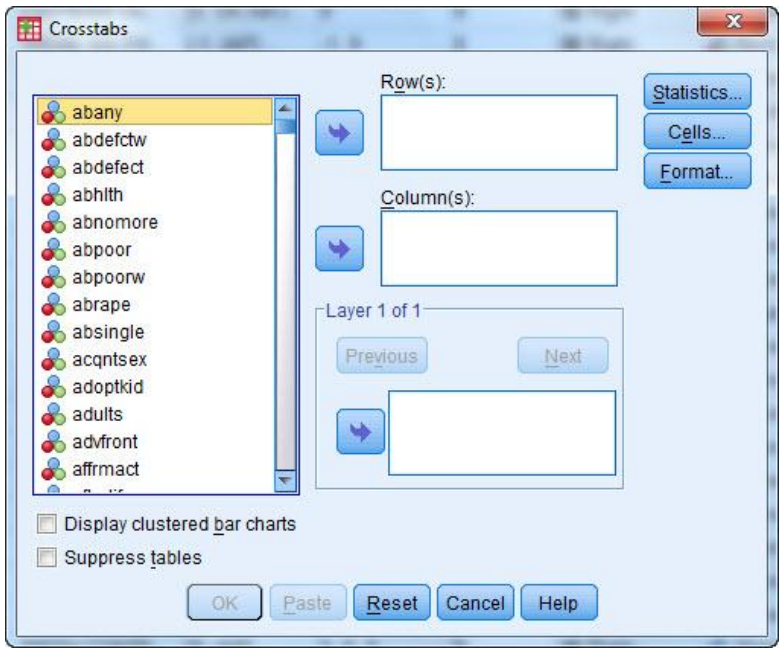

Figure 7- 4

You then choose the row (usually the dependent) variable and column (usually the independent) variable. In Appendix A, you will see that there are seven variables that deal with opinions about abortion. Let's choose "ABANY" (Abortion if a woman wants one for any reason) for our row variable and "SEX" (Respondent's sex) for column variable. To do this, select the variable you want from the list and click on it to highlight it, then use the arrow keys to the right of the list box to move the variable into either the row or the column box (for now, ignore the bottom box [layers] – more about it later). If you've done everything correctly, your screen will look like Figure 7-5. Don't click OK yet!

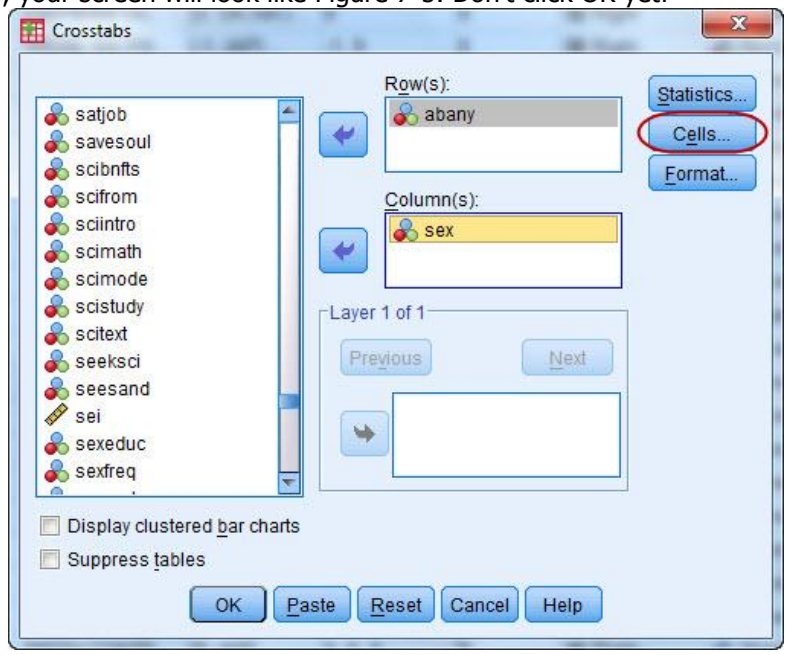

Figure 7- 5

Note the buttons at the right of the Crosstabs dialog box. Click on "Cells.." first. Here you have a number of choices for the information you would like to have in each cell of your table. The "Observed" box should already be selected – it shows the actual number of cases in each cell. You will also want to see

percentages as well as raw numbers so that you can easily compare groupings of different sizes. You should always make sure that each category of the independent variable totals 100%; our general rule is to have the dependent variables be the rows and the independent variables the columns. So choose "Columns" for the percentages as in Figure 7-6.

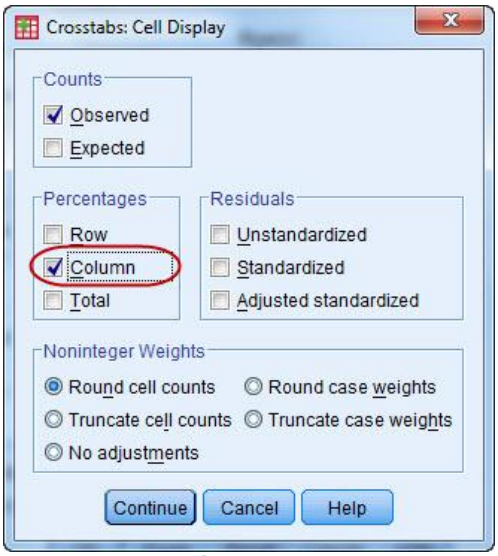

Figure 7- 6

Now click on "Continue" to get back to the Crosstabs dialog box. Once you are back there, click OK. SPSS will now open the Output Viewer window which will show you your table (see Table 7.1).

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ABORTION IF WOMAN WANTS FOR ANY REASON * RESPONDENTS SEX	1298	64.2%	725	35.8%	2023	100.0%

ABORTION IF WOMAN WANTS FOR ANY REASON * RESPONDENTS SEX Crosstabulation

			RESPONDENTS SEX		
			MALE	FEMALE	Total
ABORTION IF WOMAN WANTS FOR ANY REASON	YES	Count	262	288	550
		% within RESPONDENTS SEX	44.6%	40.6%	42.4%
	NO	Count	326	422	748
		% within RESPONDENTS SEX	55.4%	59.4%	57.6%
Total		Count	588	710	1298
		% within RESPONDENTS SEX	100.0%	100.0%	100.0%

Table 7- 1

The Case Processing Summary shows the Valid, Missing, and Total cases. The high percent of missing cases here reflects the people who were not asked this particular question in the survey. The Valid N (number of cases) is used in the table.

The Crosstabs shows the 1,298 valid cases arranged in a table that shows what percent of men and women said either Yes or No to the "ABANY question. Note that 44.6% of the men and 40.6% of the women said Yes, a percentage point difference of 4. And the exact same percentage points difference (4%) is seen between men and women who said "No". Our initial conclusion here might be that on abortion issues, there's little difference between men and women in their responses. Is this correct, or did you stop your analysis a little too soon? Let's look at a different abortion question. Repeat the steps above, but use "ABRAPE" as your dependent variable this time. Your results should look like Table 7.2

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PREGNANT AS RESULT OF RAPE * RESPONDENTS SEX	1280	63.3%	743	36.7%	2023	100.0%

PREGNANT AS RESULT OF RAPE * RESPONDENTS SEX Crosstabulation

			RESPONDENTS SEX		
			MALE	FEMALE	Total
PREGNANT AS RESULT OF RAPE	YES	Count	466	516	982
		% within RESPONDENTS SEX	80.1%	73.9%	76.7%
	NO	Count	116	182	298
		% within RESPONDENTS SEX	19.9%	26.1%	23.3%
Total		Count	582	698	1280
		% within RESPONDENTS SEX	100.0%	100.0%	100.0%

Table 7- 2

Now we see that 80% of the men and 74% of the women said Yes to "Abortion if a woman is pregnant as the result of rape." The most important thing we discover when we compare Figure 5-6 with Figure 5-7 is the large difference between total Yes answers (42.4% compared with 76.7%), which indicates that "abortion" as an issue needs to be broken down into specific conditions if you want to study it in depth. We also see that there is now a little bigger difference between men and women on this particular question. But is it a significant difference? To answer that, we will need to do some statistical analysis.

For our next Cross Tabulation, again go to the menu and choose "Analyze," "Descriptive Statistics," and "Crosstabs". In the Crosstabs dialog box place "ABRAPE" as the row variable and "SEX" as the column variable. Now click on the "Statistics.." button, then "Chi-square" to obtain a measure of statistical significance, and on "Phi and Cramer's V," which are measures of the strength of association between two variables when one or both are at the nominal level of measurement. **Phi** is appropriate for tables with two rows and two columns, while **Cramer's V** is appropriate otherwise. Your dialog box should look like Figure 7-7.

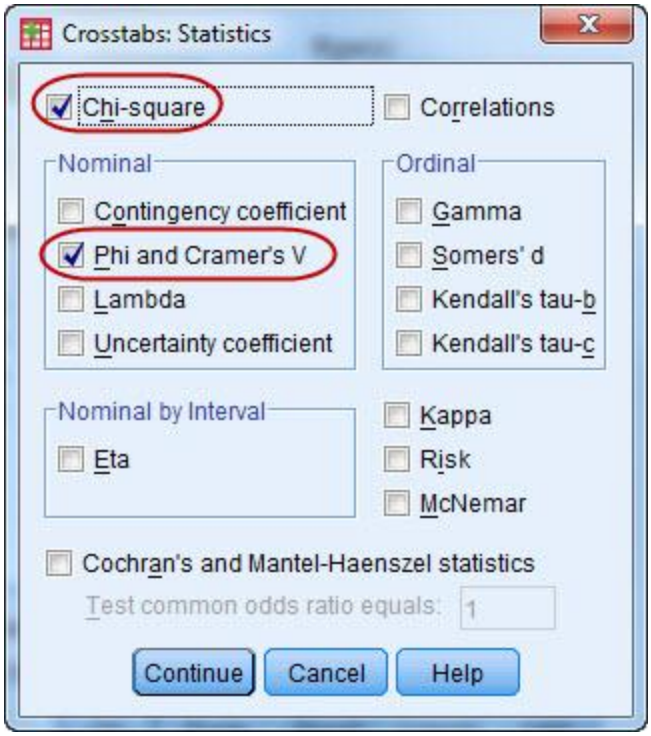

Figure 7- 7

Click on "Continue," then "OK." The table in Table 7-2 re-appears, but with some additional information (you might have to scroll down to see it) – look for "Chi-Square Tests" (Table 7-3).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.706[a]	1	.010		
Continuity Correction[b]	6.366	1	.012		
Likelihood Ratio	6.759	1	.009		
Fisher's Exact Test				.010	.006
Linear-by-Linear Association	6.701	1	.010		
N of Valid Cases	1280				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 135.50.

b. Computed only for a 2x2 table

Table 7- 3

Several different versions of chi-square (Pearson's is probably the most familiar) all indicate that the relationship in our table would occur by chance less than one time in a thousand. The Cramer's V of .080 in Figure 7-4 (Symmetric Measures), however, indicates that the relationship is not a strong one.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.072	.010
	Cramer's V	.072	.010
N of Valid Cases		1280	

Table 7- 4

Let's look at a somewhat different table. For many years, scholars have observed that in the U.S., compared to other industrialized countries, social class has relatively little impact on political attitudes and behavior. To find this out, click on "Analyze," "Descriptive Statistics," and "Crosstabs." If the variables you used before are still there, click on the "Reset" button, then move "POLVIEWS" to the "Row(s):" box and "INCCAT98" to the "Column(s):" box. Since both of these variables are ordinal, we'll want to obtain different statistics to measure their relationship. Click on "Statistics" and then on "Kendall's tau c". (**Tau c** is a measure of association that is appropriate when both variables are ordinal and do not have the same number of categories.) Click on "Continue," then on "OK." What do the results show?